OPTIMAL FIR FILTERS WITH ALMOST LINEAR PHASE

Daniel Seidner

Creo IL. Ltd.
e-mail: dani.seidner@creo.com

ABSTRACT

FIR filters are known to be stable and have a linear phase when symmetry properties. e.g., $h[n] = h[M-n]$, are kept. A common FIR filter design method is the Parks-McClellan algorithm. In this algorithm, linear phase FIR filters which are optimal in the minimax sense, are designed. Theses filters have the form of $H(\omega) = A(\omega)e^{j(\beta-\omega\alpha)}$, where $A(\omega)$ is real, α is an integer or an integer plus $1/2$ and β is 0 or $\pi/2$. These FIR filters are always symmetric or anti-symmetric. We introduce a simple procedure for designing Almost Linear Phase FIR filters, having a similar form of $\mathbf{H}(\omega)$ but an arbitrary α, that are optimal in a similar sense.

1. INTRODUCTION

Generalized Linear Phase (GLP) filters, are filters with a Discrete Time Fourier Transform (DTFT) given by

$$H(\omega) = A(\omega)e^{j(\beta-\omega\alpha)} \tag{1}$$

where $A(\omega)$ is real. It is easy to see, [1], [2], that GLP filters should satisfy

$$\sum_{n=-\infty}^{\infty} h[n]\sin(\omega(n-\alpha)+\beta) = 0 \tag{2}$$

It is easy to see that there are 4 types of causal FIR filters that satisfy equation(2), i.e., having generalized linear phase. Type-I and type-II GLP FIR filters, have symmetric impulse responses $h[n]$, i.e., $h[n] = h[M-n]$ for $n = 0, 1, \ldots, M$. The difference between the two types is that a type-I filter has an even M. i.e., an odd number of coefficients, while a type-II filter has an odd M, i.e., an even number of coefficients. Similarly, Type-III and type-IV GLP FIR filters, have anti-symmetric impulse responses, i.e., $h[n] = -h[M-n]$ for $n = 0, 1, \ldots, M$. Type-III filters have even M (and $h[M/2] = 0$), and type-IV filters have an odd M.

In all of the four types we have $\alpha = M/2$. In types I and II we have $\beta = 0$ while in types III and IV we have $\beta = \pi/2$.

Filter specifications are usually given in the frequency domain. We state the desired passband and stopband frequencies and the desired stopband and passband attenuation. A typical Low Pass Filter (LPF) specifications include therefore, ω_p, ω_s, and the amplitude $A_d(\omega)$ of the desired filter in the passband, i.e., at $\omega \in [0, \omega_p]$ and the stopband, i.e., at $\omega \in [\omega_s, \pi]$. For a LPF we would like to have

$$A_d(\omega) = \begin{cases} 1 & \omega \in [0, \omega_p] \\ 0 & \omega \in [\omega_s, \pi] \end{cases} \tag{3}$$

Usually when FIR filters are designed, they are designed to be one of the four types we mentioned above. Thus, $H_d(\omega)$ includes a phase factor of $e^{j(\beta-\omega\alpha)}$, where $\alpha = M/2$ and $\beta = 0$ or $\pi/2$. For LPF, we therefore have

$$H_d(\omega) = A_d(\omega)e^{-j\omega M/2} \tag{4}$$

However, the actual filter we acquire, having $M+1$ coefficients, i.e., $h[n]$ for $n = 0, 1, \ldots, M$, does not satisfy equation(4). The filter's actual frequency response, denoted $H(\omega)$, is not equal to the desired $H_d(\omega)$. The error in $H(\omega)$ is given by

$$E(\omega) = H_d(\omega) - H(\omega) \tag{5}$$

A common approach is to use the minimax (Chebyshev) criterion, i.e., to seek for the filter $h[n]$ having $H(\omega)$ that minimizes the maximal error in the interval $I = \{\omega \in [0, \omega_p]$ or $\omega \in [\omega_s, \pi]\}$. Thus we seek $h[n]$ having $M+1$ coefficients (for $n = 0, 1, \ldots, M$) producing the minimal weighted error δ given by

$$\delta = \min_{h[n]} \max_{\omega \in I} \{W(\omega)|E(\omega)|\} \tag{6}$$

where $W(\omega)$ is some weighting function. Thus, the optimal approximated filter $h[n]$ satisfies

$$h[n] = \arg\min_{h[n]} \max_{\omega \in I} \{W(\omega)|E(\omega)|\} \tag{7}$$

The Parks-McClellan algorithm is probably the most popular algorithm for designing minimax optimal GLP FIR filters. In this algorithm, we find the filter that minimizes δ for a given number of coefficients and a given $A(\omega)$ and $W(\omega)$.

2. ALMOST LINEAR PHASE FIR FILTERS

For an even M, we get a linear phase filter having a delay of an integer number of samples. For an odd M, we get a filter having a delay of an integer number of samples plus half a sample. Sometimes, e.g., when shifting or scaling an image, it is desired to have a delay that is not an integer or an integer plus half number of samples. The regular Park-McClellan algorithm does not take the desired phase into account. Well, to be more precise, it assumes that $h[n]$ is symmetric or anti-symmetric and therefore includes a phase factor of $e^{j\omega M/2}$. Therefore, equations (6) and (7) really mean

$$\delta = \min_{h[n]} \max_{\omega \in I} \{W(\omega)|A_d(\omega)e^{-j\omega M/2} - A(\omega)e^{-j\omega M/2}|\} \tag{8}$$

and so, we can ignore the phase factor of $e^{-j\omega M/2}$ and concentrate on

$$\delta = \min_{h[n]} \max_{\omega \in I} \{W(\omega)|A_d(\omega) - A(\omega)|\} \tag{9}$$

where $A(\omega) = H(\omega)e^{j\omega M/2}$.

The Parks-McClellan algorithm is using the fact that each of the four GLP FIR filter types can be written as $H(\omega) = e^{j(\beta-\omega M/2)}Q(\omega)\sum_{k=0}^{K} g[k]\cos(\omega k)$ and so the algorithm is searching for $A(\omega) = Q(\omega)\sum_{k=0}^{K} g[k]\cos(\omega k)$ satisfying equation (9).

For producing FIR optimal approximation (in the minimax sense) to a filter having an arbitrary delay, we need a somewhat different minimax criterion. First, we define the desired $H_d(\omega)$ as

$$H_d(\omega) = A_d(\omega)e^{-j\omega(M/2+\alpha)} \tag{10}$$

where here α is the extra delay (added to the $M/2$ "normal" delay of the GLP FIR filters). Typical values for α could be between -0.5 to 0.5, i.e., α can be a fractional delay. Then we define the minimax criterion. A possible criterion is to use an equation similar to equation (6) or (8) with the additional phase α:

$$\delta = \min_{h[n]} \max_{\omega \in I} \{W(\omega)|A_d(\omega)e^{-j\omega(M/2+\alpha)} - H(\omega)|\} \tag{11}$$

This is not so easy to achieve. However, changing the criterion to two parts as in the two following equations

$$\delta = \min_{h[n]} \max_{\omega \in I} \{W(\omega)|\mathrm{real}\{A_d(\omega)e^{-j\omega\alpha} - e^{j\omega M/2}H(\omega)\}|\} \tag{12}$$

and also

$$\delta = \min_{h[n]} \max_{\omega \in I} \{W(\omega)|\mathrm{imag}\{A_d(\omega)e^{-j\omega\alpha} - e^{j\omega M/2}H(\omega)\}|\} \tag{13}$$

makes it very easy to implement. The difference between the criterion of equation (11) to the criterion of equations (12) and (13) is in the shape of the range of the allowed error. If we consider $H_d(\omega)$ as a line in a 3 dimensional space, where the axes are ω, and the real and imaginary components of $H_d(\omega)$, then equation (11) describes a round "tube" with a radius of $\delta/W(\omega)$ around that line. That tube is the volume in which we allow $H(\omega)$ to be. If $H(\omega)$ is in that "tube", the weighted error $W(\omega)|A(\omega)e^{-j\omega(M/2+\alpha)} - H(\omega)|$ is smaller than or equal to δ. Equations (12) and (13) describe a similar volume having a square shaped cross section instead of the round cross section of equation (11).

Splitting the original criterion to two parts, the real part and the imaginary part, immediately brings us to a simple procedure for computing the approximated filter $h[n]$. Since we know from the symmetry properties of the DTFT that the real part of $H(\omega)$ is the DTFT of $h_e[n]$, where $h_e[n]$ is the even part of $h[n]$ given by

$$h_e[n] = (h[n] + h[M-n])/2 \tag{14}$$

and that imaginary part of $H(\omega)$ is the DTFT of $h_o[n]$, where $h_o[n]$ is the odd part of $h[n]$ given by

$$h_e[n] = (h[n] - h[M-n])/2 \tag{15}$$

we can design the even part $h_e[n]$ according to equation (12), and the odd part $h_o[n]$ according to equation (13) separately. Looking carefully on equations (8), (9) and the Parks-McClellan algorithm, we conclude that we should use the Park-McClellan algorithm with new specifications as in equations (16) and (17) below:

$$\delta = \min_{h_e[n]} \max_{\omega \in I} \{W(\omega)|A_d(\omega)\cos(\omega\alpha) - A_e(\omega)|\} \tag{16}$$

$$\delta = \min_{h_o[n]} \max_{\omega \in I} \{W(\omega)|A_d(\omega)\sin(\omega\alpha) - A_o(\omega)|\} \tag{17}$$

where the algorithm is searching for a symmetric (type-I or II) filter according to equation (16) and for an anti-symmetric (type-III or IV) filter according to equation (17). $A_e(\omega)$ is the amplitude response of the symmetric filter, without the $e^{-j\omega M/2}$ phase factor, i.e.,

$$H_e(\omega) = A_e(\omega)e^{-j\omega M/2} \tag{18}$$

Similarly, $A_o(\omega)$ is the amplitude response of the anti-symmetric filter, without the $e^{-j(\omega M/2-\pi/2)}$ phase factor, i.e.,

$$H_o(\omega) = jA_o(\omega)e^{-j\omega M/2} \tag{19}$$

The final filter in the frequency domain is given by

$$H(\omega) = H_e(\omega) + H_o(\omega) \tag{20}$$

and in the time domain by

$$h[n] = h_e[n] + h_o[n] \quad (21)$$

The final filter does not have a linear phase. However, its phase is almost linear and so we call it an Almost Linear Phase (ALP) FIR filter. We can increase the number of coefficients and change the weighting function in order to attenuate the deviation from linear phase. Note that for a given number of coefficients, the resulting filter is optimal in the sense of equations (12) and (13).

3. EXAMPLES

We give below two examples of ALP filters having 6 coefficients. The desired $A_d(\omega)$ of the filters is the amplitude of the frequency response of a given filter
$h_{ref}[n] = [-0.00, -0.16, 0.66, 0.66, -0.16, -0.00]$, and is shown in figure 1. In that figure we also see the actual $|H(\omega)|$ of two ALP filters designed to be optimal and have delays of $\alpha = -0.125$ and $\alpha = -0.25$. These filter were found using an exhaustive search with a relatively low resolution of 0.005 in the coefficients values without any weighting function for $h_e[n]$ and with a weighting function of 50 for $\omega \in [0, 0.8\pi]$, decreasing linearly to 10 at $\omega = 0.9\pi$ and further to 0 at $\omega = \pi$, for $h_o[n]$. We also forced $H_e(0)$ to be 1, i.e., $\sum_{n=0}^{M} h_e[n] = 1$. For $\alpha = -0.125$ we found the values of:
$h_e[n] = [-0.005, -0.150, 0.655, 0.655, -0.150, -0.005]$, $h_o[n] = [-0.020, 0.025, 0.125, -0.125, -0.025, 0.020]$ and so $h[n] = [-0.025, -0.125, 0.780, 0.530, -0.175, 0.015]$.
For $\alpha = -0.25$ we found the values of:
$h_e[n] = [-0.015, -0.115, 0.630, 0.630, -0.115, -0.015]$, $h_o[n] = [-0.040, 0.050, 0.235, -0.235, -0.050, 0.040]$ and so $h[n] = [-0.055, -0.065, 0.865, 0.395, -0.165, 0.025]$.

The resulting phase of the two filters, compared to the desired phase is depicted in figure 2.

The group delay, compared to the desired group delay, i.e., $M/2 + \alpha$, is shown in figure 3.

We see that the magnitude of the acquired filters is pretty close to the desired filter. We also see that the resulting phase is almost linear with alternations of the group delay near the desired value along most of the frequency axis.

4. CONCLUSION

In this paper we had shown a simple procedure for designing minimax optimal Almost Linear Phase FIR filters having an arbitrary fractional delay. The procedure is using the regular design methods for designing Generalized Linear Phase FIR filters. It is based on minimax criterion which is applied separately to the real and imaginary parts of the desired filter in the frequency domain, thus, producing the even and odd parts of the approximated filter separately.

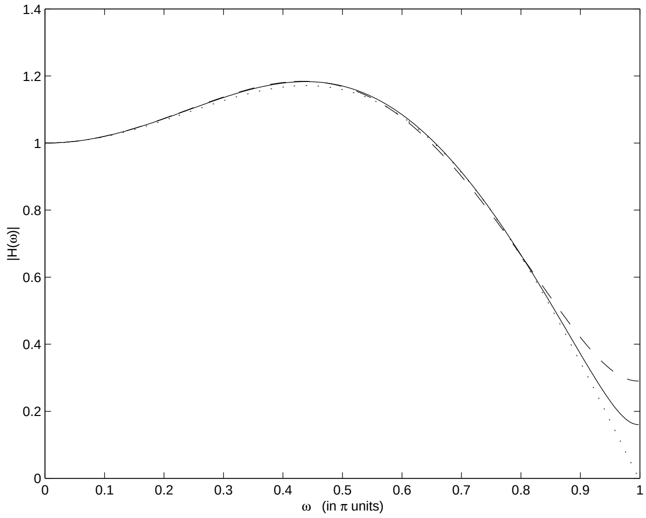

Fig. 1. $|H_d(\omega)|$ (dotted line) vs. $|H(\omega)|$ for $\alpha = -0.125$ (solid line) and for $\alpha = -0.25$ (dashed line)

5. REFERENCES

[1] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*, Prentice Hall, Englwood Cliffs, New Jersey, 1999.

[2] L. R. Rabiner and B. Gold. *Theory and Application of Digital Signal Processing*, Prentice Hall, Englwood Cliffs, New Jersey, 1975.

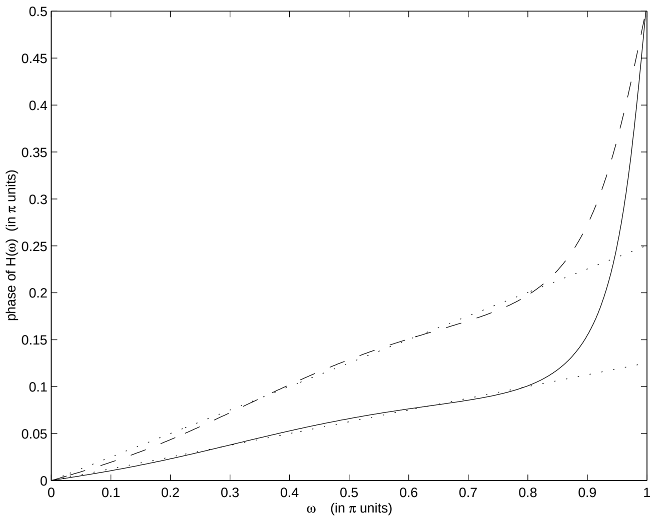

Fig. 2. Phase of $H(\omega)$ for $\alpha = -0.125$ (solid line) and for $\alpha = -0.25$ (dashed line) vs. the desired phase (dotted lines)

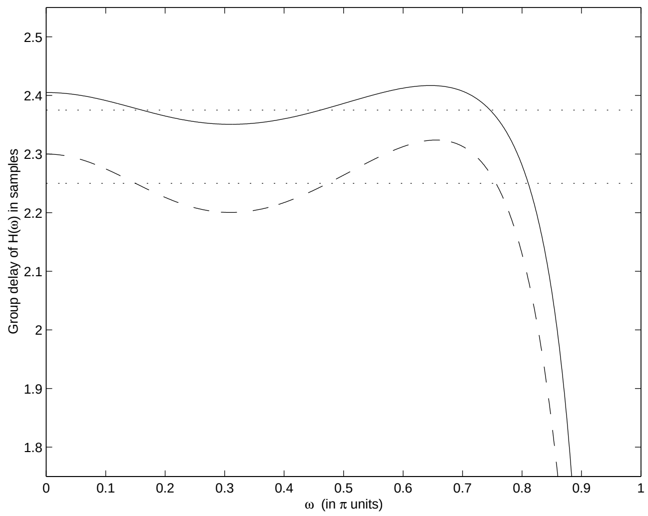

Fig. 3. The group delay of $H(\omega)$ for $\alpha = -0.125$ (solid line) and for $\alpha = -0.25$ (dashed line) vs. the desired delay (dotted lines)